

# Anonymous Real-Time Speech (ARTS) Proposers' Day

Mark Becker | Program Manager | 2023 June 27



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology



# Welcome to the ARTS Proposers' Day



Thank you for your interest in this program and participating in this event

To assure a clear broadcast stream, audio and video are disabled for meeting participants

Comments and questions can be submitted to the IARPA team via the WebEx chat tool submission or via index cards for in-person attendees

- Please direct questions to “All Panelists” in the chat if you are virtual

Questions submitted to the alias ([dni-iarpa-ARTS-proposers-day@iarpa.gov](mailto:dni-iarpa-ARTS-proposers-day@iarpa.gov)) prior to this meeting and during this presentation, and corresponding answers, may be posted in writing online



# Disclaimers



- This presentation is provided solely for information and planning purposes
- The Proposers' Day does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers' Day changes the requirements set forth in a BAA
- **The BAA language supersedes anything presented or said by IARPA at the Proposers' Day**
- This meeting is being recorded and will be posted for public viewing
- For those viewing the recording, email aliases and POCs may be dated, please refer to [IARPA.gov](http://IARPA.gov) for updated information.

# Proposers' Day Goals

1. Familiarize participants with IARPA's interest in the ARTS program and solicit questions and feedback
2. Foster discussion of complementary capabilities among potential program participants, i.e., TEAMING
  - Teaming information can be found at the following address:  
<https://www.iarpa.gov/index.php/research-programs/arts>
  - An attendance list, with contact information of participants who approved of sharing will be distributed soon
  - The chat feature is enabled for participants to plan future discussions associated with teaming
  - Teaming interests, capability summaries, and lightning talk slides will be posted publicly on the IARPA ARTS webpage until the BAA submission period closes

Please ask questions and provide feedback, this is your chance to alter the course of events.  
Please talk with others, find great team members.



# Feedback and Questions



- Questions can be submitted until 11:00am ET.
- There will be a break after the contracting presentation at 11:00am ET.
- Responses to selected questions will be broadcast at 12:30pm ET, so please don't log out or close your WebEx connection.
  - All programmatic and contractual questions will be captured but will not be answered in this session
- Feedback (but not questions) about the draft technical section may be submitted to the IARPA team email at [dni-iarpa-ARTS-proposersday@iarpa.gov](mailto:dni-iarpa-ARTS-proposersday@iarpa.gov).
  - A new alias will be established for when the full BAA is released
- After this Proposers' Day, IARPA will review all the feedback received for a final BAA to be posted on SAM.gov and Grants.gov.



# Teaming



- Participants are encouraged to find partners and collaborators . . . someone might have a missing piece of your puzzle.
- Lightning talks will take place following the Program presentations.
- Collaborating and capability summaries will be accepted, with minimal review for appropriateness, and made available to the public.
  - Teaming documents and summaries can be submitted until the BAA closes, submit to [dni-iarpa-ARTS-proposersday@iarpa.gov](mailto:dni-iarpa-ARTS-proposersday@iarpa.gov).
  - If you would prefer your information not be shared (any recorded videos cannot be modified or removed) email [dni-iarpa-ARTS-proposersday@iarpa.gov](mailto:dni-iarpa-ARTS-proposersday@iarpa.gov).



# Agenda



Time	Topic	Speaker
9:30am-9:40am	Welcome, Logistics, Proposers' Day Goals	Mark Becker, Program Manager
9:40am-9:50am	IARPA Overview	Robert Rahmer, Director Office of Analysis Research, IARPA
9:50am-10:40am	ARTS Program Overview	Mark Becker
10:40am-11:00am	Contracting Overview	Kevin Woody, Department of Interior
11:00am-12:30pm	<b>Break (Submit questions in chat or drop boxes before 11:00am)</b>	
12:30pm-1:30pm	Answers to Selected Technical Questions	Mark Becker
1:30pm-1:35pm	Introductions to Lightning Talks	Mark Becker
1:35pm-3:45pm (est.)	Lightning Talks*	Potential Performers
3:45pm-5:00pm	Informal Teaming Discussions*	In-Person Participants

\*The Government will not attend these events





# LIGHTNING TALKS AGENDA



Time	Speaker	Institution	In person
1:35pm-1:45pm	Dr. Tor Finseth	Honeywell	YES
1:45pm-1:55pm	Dr. Jonathan Wintrode	Raytheon AST	YES
1:55pm-2:05pm	Yishay Carmiel	Meaning	YES
2:05pm-2:15pm	Dr. Kemal Davasliouglu	Nextcepta	YES
2:15pm-2:25pm	Prof. John Hansen	UT Dallas	YES
2:25pm-2:35pm		Break	





# LIGHTNING TALKS AGENDA CONTINUED



Time	Speaker	Institution	In person
2:35pm-2:45pm	Dr. Erica Gold	CSU San Marcos	NO
2:45pm-2:55pm	Dr. Jim Schroeder	Florida R&D	NO
2:55pm-3:05pm	Dr. Scott Grigsby	PAR Government	NO
3:05pm-3:15pm	Jean-Luc Crebouw	CandyVoice	NO
3:15pm-3:25pm	Dr. Ritu Chadha	Peraton Labs	NO
3:25pm-3:35pm	Dr. Homayoon Beigi	Recognition Technologies	NO
3:35pm-3:45pm	Dr. Nirupam Roy	UMD College Park	NO
3:45pm-5:00pm	Informal Teaming Discussions		In-Person Participants



# IARPA Overview

Robert Rahmer | Director, IARPA Office of Analysis | ARTS Proposers' Day, 27 June 2023



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

# Office of the Director of National Intelligence





# IARPA Mission

IARPA envisions and leads *high-risk, high-payoff research* that delivers innovative technology *for future overwhelming intelligence advantage*

- Our problems are **complex** and **multidisciplinary**
- We emphasize **technical excellence** & **technical truth**

- **Bring the best minds to bear on our problems**
  - Full and open competition to the greatest possible extent
  - World-class, term-limited Program Managers
- **Define and execute research programs that:**
  - Have goals that are clear, ambitious, credible and measurable
  - Run from three to five years
  - Publish peer-reviewed results and data, to the greatest possible extent
  - Employ independent and rigorous Test & Evaluation
  - Involve IC partners from start to finish
  - Transition new capabilities to intelligence community partners

- Technical and programmatic excellence are required
- Each program has a clearly defined and measurable end-goal
  - Intermediate milestones to measure progress are also required
  - Every program has a beginning and an end
- This approach, coupled with term-limited PM positions, ensures
  - IARPA does not “institutionalize” programs
  - Fresh ideas and perspectives are always coming in
  - Status quo is always questioned
  - Only the best ideas are pursued, and only the best performers are funded





# IARPA Snapshot



IARPA's research portfolio is diverse, including math, physics, chemistry, biology, microelectronics, neuroscience, linguistics, political science, cognitive psychology, and more.

- 70% of completed research transitions to U.S. Government partners
- 3,000+ journal articles published
- IARPA funded researchers have been awarded the **Nobel Prize in Physics** for quantum computing research, a **MacArthur Fellowship**, and a **Bell prize**
- IARPA serves on National Science and Technology Council (NSTC) committees and actively engages with the White House BRAIN Initiative, National Strategic Computing Initiative, and the NSTC Select Committee on Artificial Intelligence, the NSTC Subcommittee on Quantum Information Science (SCQIS), and NSTC Subcommittee on Economic and Security Implications of Quantum Science (ESIX)



# ENGAGE WITH US

Throughout our website you can learn more about engaging with us on our highly innovative work that is having a positive impact in the Intelligence Community and society in general. Click on any of the below links to learn more.

[iarpa.gov](http://iarpa.gov) | 301-243-1995

[dni-iarpa-info@iarpa.gov](mailto:dni-iarpa-info@iarpa.gov)

- Reach out to our Program Managers.
- Schedule a visit if you are in the DC area or invite us to visit you



## Open BAAs

Broad Agency Announcements (BAAs) solicit research proposals for specific programs. Learn more about current BAA opportunities and ways to get involved...



## Requests For Information

Requests for Information (RFIs) are designed to gather more information on an idea in an area in which our program managers are not fully informed...



## Seedlings

Seedlings are typically 9 - 12 month research efforts that are less than \$1M in cost. They are intended to address highly innovative ideas and concepts within...

# ARTS Overview



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

# Anonymous Real-Time Speech (ARTS)

The ARTS program will develop new algorithms for real-time voice modification that defeat a wide range of threats. An ARTS system will:

1. Decrease effectiveness of automated speaker identification (SID) tools.
2. Reduce utility of Forensic Speech Analysis (FSA) used to assess static traits
3. Reduce the ability to apply AI/ML techniques to determine dynamic traits.
4. Perform in real-time.
5. Be understandable by automatic speech recognition (ASR) tools
6. Sound natural according to synthetic speech detection (SSD) methods.



# The Speech Privacy Problem

Microphones are ubiquitous and speech is recorded everywhere

- Voice input tools
- Call centers
- Electronic devices
- Eavesdropping

An individual's speech can **reveal their identity** through speaker identification (SID)

Forensic speech analysis (FSA) can be used to **profile static traits of a speaker**

- Age
- Educational background
- Ethnicity/Nationality
- Geographical origin
- Dialects
- Gender

FSA can be used to **infer dynamic states of a speaker**

- Stress
- Intoxication
- Physical health
- Fear

The ARTS program will develop methods to modify **live** conversational speech so that personally identifiable information is concealed, but other aspects of speech (**understandability, naturalness**) are preserved.

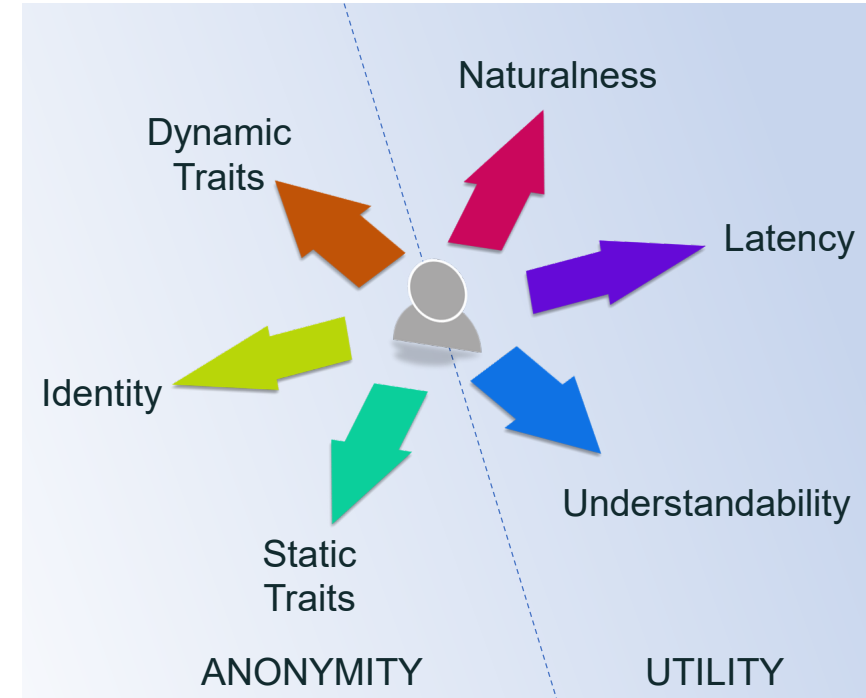
# The ARTS Challenge

The ARTS Program strives for a novel technologies to address competing needs of **anonymity** and **utility**. Whereas current research focuses on a small subset of these areas, ARTS is looking at six different dimensions simultaneously.

- In addition to the SID threat, ARTS is tackling the threat of forensic analysis. ARTS will change *how* things are said.
- Constraints such as real-time, accuracy, and realism compete with methods to anonymize speech, and ARTS will discover how one requirement impacts another.

ARTS is a privacy enhancing technology that will use the threat model of an informed attacker, who knows the anonymization - methods should rely on the strength of the algorithms.

ARTS will start with English, but over the course of the program it will take on a multi-lingual component to extend capabilities other languages.





ARTS will produce systems that will:

- **TA1:** De-identify speech - SID scores from transformed test segments should be statistically indistinguishable across different targets enrolled in the database.
- **TA2:** Remove existing static traits and replace them with pre-selected ones - Panels of human evaluators should believe speakers have the pre-selected traits.
- **TA3:** Modify speech so that statistical features related to emotions are eliminated - AI/ML techniques used to assess short term states will be inconclusive.

# Utility Constraints

The ARTS Program has three main utility constraints to ensure the ARTS system is useable:

- **Latency** – the system should perform in real-time so that it is suitable for conversational speech.
- **Understandability** – transformed speech should be as accurate as the original speech.
- **Naturalness** – speech should sound like it is spoken by a human.



# Technical Areas and Utility Constraints



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

- SPEECH CORPORA
- TARGET PROFILES
- SID APPROACH
- SYSTEM REQUIREMENTS

## ARTS

- TA1: Perform Speaker De-Identification
- TA2: Remove actual static traits, replace with target traits
- TA3: Remove dynamic emotional traits

Utility Requirements:

1. Produce speech output within a specified latency
2. Process speech so it is no less understandable
3. Create output speech that sounds natural

- ## SPEECH PROCESSING ALGORITHMS
- Module 1: SDID
  - Module 2: Static trait substitution
  - Module 3: Dynamic trait removal



# TA1: Speaker De-Identification

- SPEECH CORPORA
- TARGET PROFILES
- SID APPROACH
- SYSTEM REQUIREMENTS

## ARTS

**TA1: Perform Speaker De-Identification**

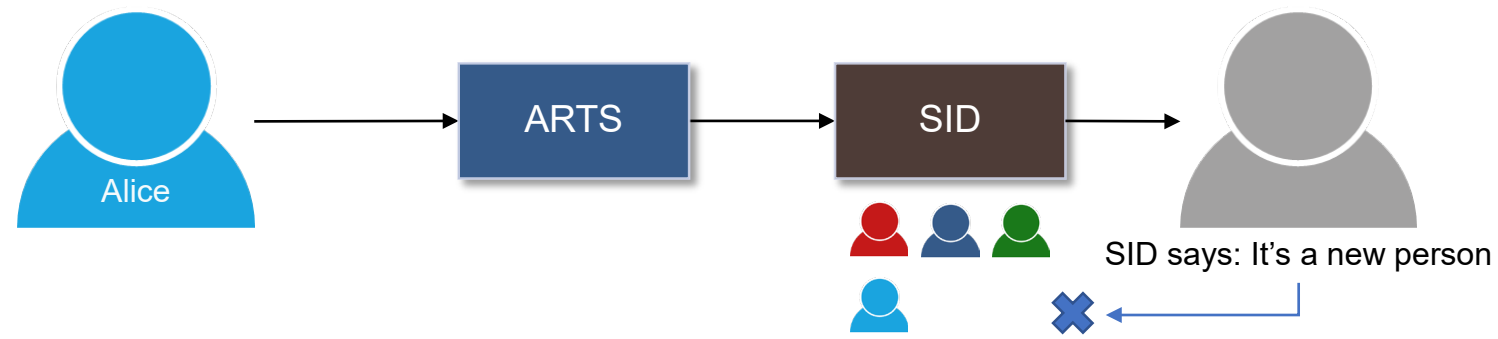
TA2: Remove actual static traits, replace with target traits

TA3: Remove dynamic emotional traits

Utility Requirements:

1. Produce speech output within a specified latency
2. Process speech so it is no less understandable
3. Create output speech that sounds natural

- ## SPEECH PROCESSING ALGORITHMS
- Module 1: SDID
  - Module 2: Static trait substitution
  - Module 3: Dynamic trait removal



# TA1: Innovations for Defeating SID

Performers will research and develop a system to modify speech so that the output speech will defeat SID systems, agnostic to the methods used in the SID attack

- An eavesdropper Eve has full access to the algorithms used to anonymize speech.
- Speakers of test segments are enrolled in the SID database.
- Parameters and other data are considered secret.

Performers will devise these algorithms, so they transform speech within specified latency requirements.

The output speech will be understandable and sound natural.

## Innovations that will help defeat SID

## Remarks

Advanced low-level representations of speech, such as i-vectors, x-vectors, often used in VC techniques

Speaker recognition tasks and VC technologies depend on these representations.

New, fast speech synthesis models that create realistic sounding artificial speech.

Commercial technologies such as VALL-E and Resemble.ai that create realistic, emotional speech from text.

Text-free speech synthesis

Current research trend in speech synthesis that transforms speech without the need for text or symbolic representations, speeding up the process.

- SPEECH CORPORA
- TARGET PROFILES
- SID APPROACH
- SYSTEM REQUIREMENTS

### ARTS

TA1: Perform Speaker De-Identification

TA2: Remove actual static traits, replace with target traits

TA3: Remove dynamic emotional traits

Utility Requirements:

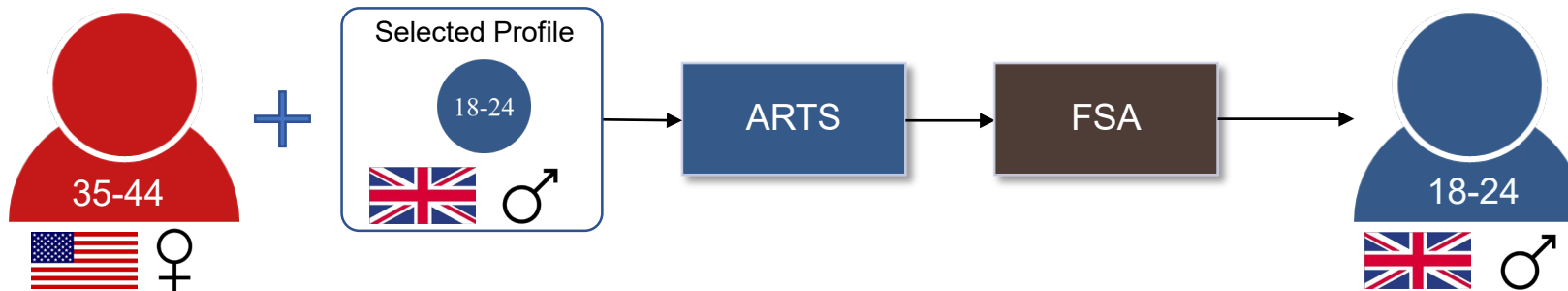
1. Produce speech output within a specified latency
2. Process speech so it is no less understandable
3. Create output speech that sounds natural

### SPEECH PROCESSING ALGORITHMS

Module 1: SDID

Module 2: Static trait substitution

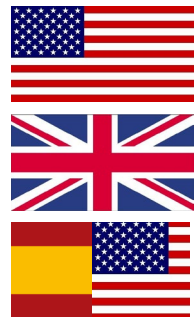
Module 3: Dynamic trait removal



# TA2: Innovations for Defeating FSA (Static Traits)

The performers will transform speech to remove static traits and replace them with pre-selected ones.

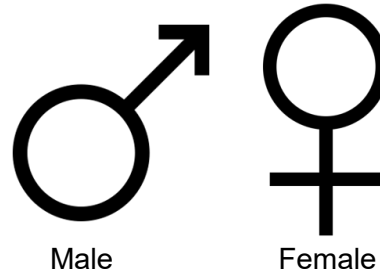
- Metadata associated with speech segments will include profile information
- Target pseudo profiles will be selected randomly.
- The goal is to change speech so that a listener assess each category to match the pseudo profile.



Standard American English

Standard Southern British English

Language 1 (L1): Spanish  
Language 2 (L2): English



Male

Female



Young



Adult



Senior

## Innovations that will help defeat FSA for static traits

## Remarks

Foreign accent conversion methods modifying frequency, aperiodicity, phonetic information.

Academia and industry research lead research for language learning and communication with call centers outsourced overseas.

Detection, analysis, and understanding of speech disfluencies.

Research from speech-to-text systems that have improved detection and removal of disfluencies.



# TA3: FSA and Dynamic Traits

- SPEECH CORPORA
- TARGET PROFILES
- SID APPROACH
- SYSTEM REQUIREMENTS

## ARTS

TA1: Perform Speaker De-Identification

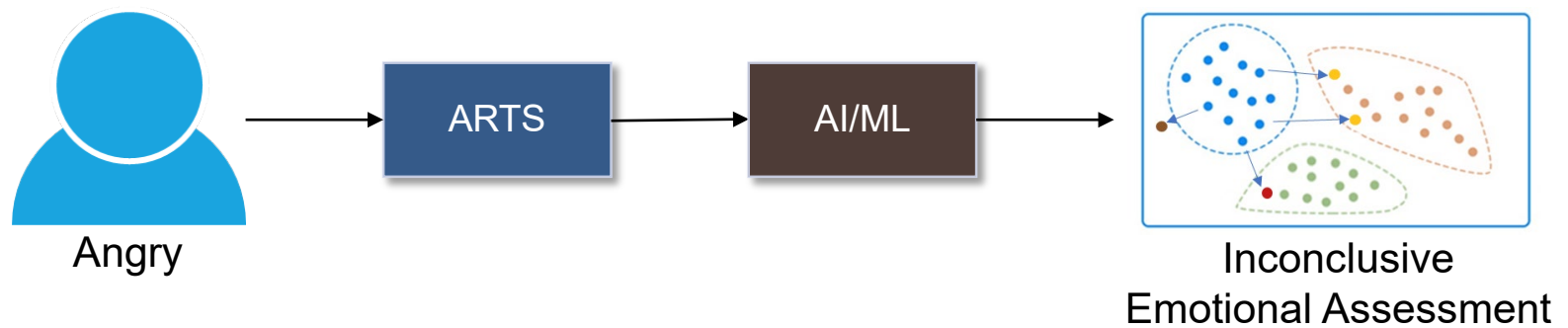
TA2: Remove actual static traits, replace with target traits

TA3: Remove dynamic emotional traits

Utility Requirements:

1. Produce speech output within a specified latency
2. Process speech so it is no less understandable
3. Create output speech that sounds natural

- ## SPEECH PROCESSING ALGORITHMS
- Module 1: SDID
  - Module 2: Static trait substitution
  - Module 3: Dynamic trait removal





# TA3: Innovations for Defeating FSA (Dynamic Traits)

The performers will transform speech to remove dynamic traits.

- Existing data sets for speech emotion recognition contain categorized segments of emotion or other short-term states.
- Low-level descriptor (LLD) features for energy, spectral components, and voicing, are often used to classify emotion based on speech.
- The goal will be to alter speech so that assessment of dynamic traits is no better than a guess.

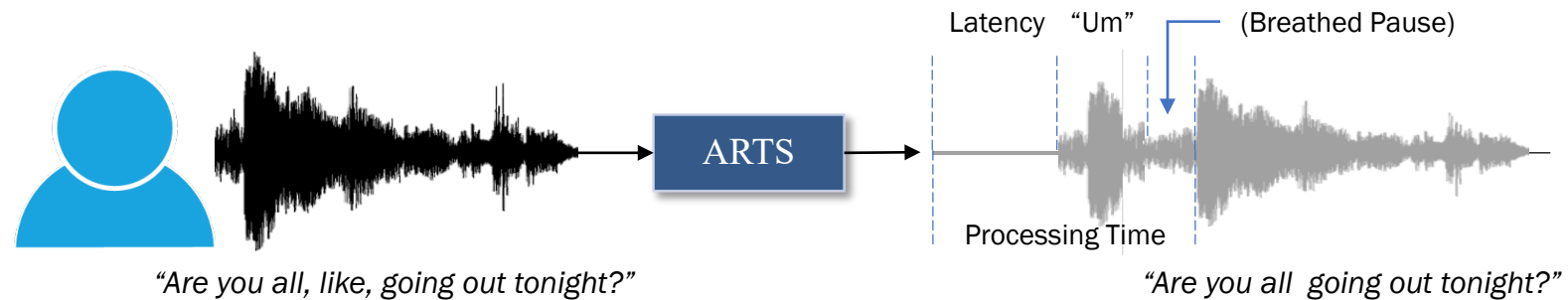
Innovations that will help defeat FSA for dynamic traits	Remarks
Advances in Generative Adversarial Networks	GANs have played a significant role in adjacent fields of speech enhancement, and have been shown to be well-suited for general speech processing.
Techniques for source modeling of emotional speech	Speech emotion recognition (SER) systems have advanced tremendously for use in modern conversational systems, such as Google Assistant, Siri, and Alexa.
Innovations from Universal Adversarial Perturbations (UAP)	Competitions at International Conference on Acoustics, Speech, and Signal Processing aimed to defeat DNNs that use high-dimensional speech features.

# Utility Constraint: Latency - What Real-Time Means

Real-time speech transformation means the duration that elapses between speech input speech and voice activity detection (VAD).

- It does not mean delay between matching words.
- Spontaneous speech often has delays, but people manage this naturally in conversational speech.
- For instance, use of discourse markers can buy time while formulating the main message.
- Real-time is crucial for conversational speech, there is a limit to how much latency people will accept.
- ARTS systems will have latency on the order of common communication channels.

Channel	Acceptable Latency
Satellite	450-550 ms
VoIP	20-300 ms
Cell Phone	30-100 ms
Fiber	3.33 ms

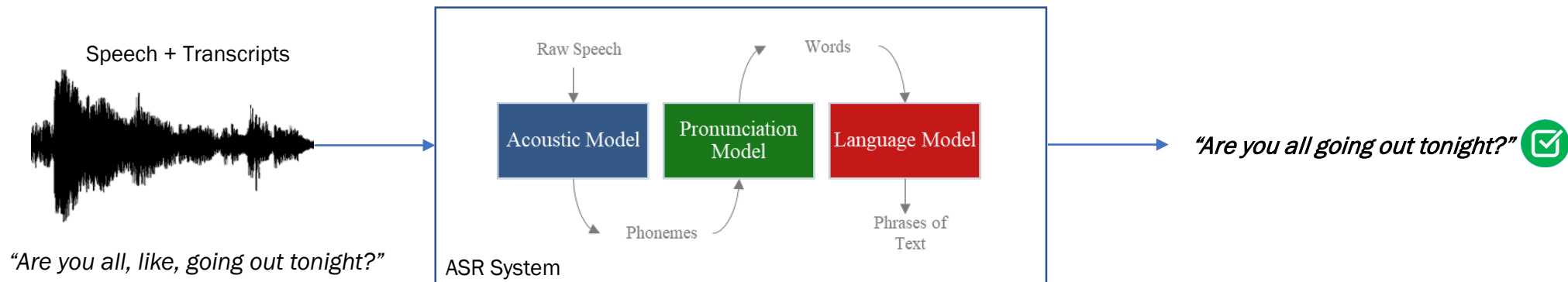


Transforming speech can result in sounds that are difficult to understand.

- For instance, interviews that distort the voice from an anonymous source often require subtitles for understandability.

ASR systems (speech-to-text), are reliable methods to assess how well speech can be understood.

- Annotated speech data sets allow ASR systems to set a baseline for understandability.



Despite huge technological advances in speech synthesis, people are still good at assessing whether speech is from a human or computer.

ARTS seeks innovations to sanitize speech characteristics that can be used to identify or label an individual. This includes:

- **Phonetic (Segmental)**
  - Trap vowels vs. cardinal vowels (/æ/ vs. /a/)
  - Glottal stops (“butter” vs. “budder”)
  - Word final rhoticity (standard vs. received)
- **Phonetic (Suprasegmental)**
  - Prosody
    - Fundamental frequency
    - Phonations (vocal fry)
    - Intonations (high rising terminal, fall-rise intonation)
    - Articulation rate and speech rate
- **Non-Phonetic (Linguistic)**
  - Fillers (“uh”, “um”, “like”, “you know”, “I mean”)
  - Restarts (repetitions, corrections, restarts)
- **Non-Phonetic (Non-Linguistic)**
  - Pauses (silent, breathed, filled)
- **Non-Phonetic (Lexical Choices)**
  - Keywords/slang

# ARTS Program Phases



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

Phase 1 will focus on English

- Initial data will consist of Standard American English (SAE) and Standard Southern British English (SSBE) data.
- Final exams will include L2 English, with L1=Spanish to push into Phase 2.

Phase 2 will begin with Spanish

- Initial data will consist of dialects from Spain
- Additional data will use L2 Spanish, with L1=English
- Stretch goals will train/test on Russian, Arabic, and Chinese

### Phase 1

	Standard American English
	Standard Southern British English
	Language 1 (L1): Spanish Language 2 (L2): English

### Phase 2

	North Central Peninsular Spanish		Russian
	Eastern Andalusian Spanish		Modern Standard Arabic
	L1: English L2: Spanish		Mandarin Chinese



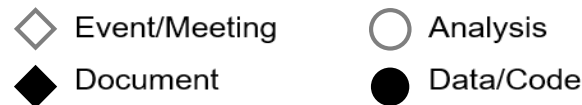
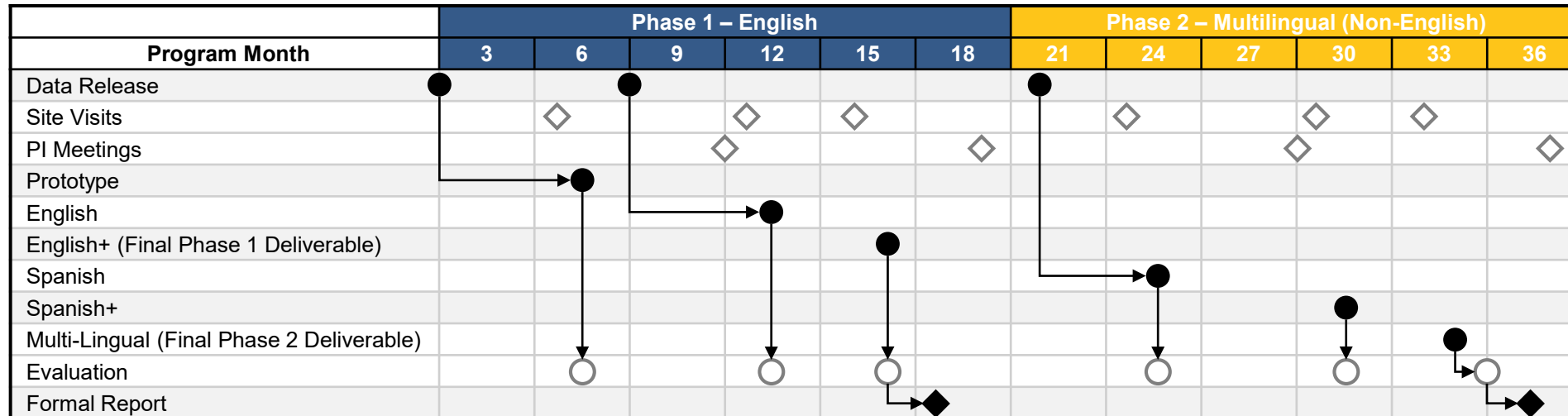
# Program Phases

ARTS aims to satisfy customers with early and continuous delivery - interim software deliverables required approximately every 6 months.

- Software will adhere to an API developed for ARTS and must be trainable by T&E.

Phase 1 will focus on English, with final exams consisting of American, British, and Accented dialects.

Phase 2 will start with Spanish consisting of Castilian, Andalusian, and Accented dialects, with stretch goals of Russian, Arabic, and Mandarin Chinese





# ARTS Phases

The ARTS Program will consist of two 18-month phases.

Program Phase	Languages or Dialects	Data Provided (Program Month)	System Delivery (Program Month)	User Knows SID Attack Method	Attacker Knows SDID Method	Areas Evaluated
Phase 1 (English)	Regional US	0	5	YES	NO	TA1, Latency, Understandability
	American, British	6	11	NO	NO	TA1, TA3, Latency, Understandability
	American, British, Accented	N/A	16	NO	YES	All
Phase 2 (Multilingual)	Castilian, Andalusian	19	23	YES	NO	TA1, Latency, Understandability
	Castilian, Andalusian, Accented Spanish	N/A	28	NO	NO	TA1, TA3, Latency, Understandability
	Arabic, Russian, Mandarin Chinese	N/A	33	NO	YES	All

# Evaluation and Metrics



Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

Robust, independent test and evaluation is a crucial part of every IARPA program

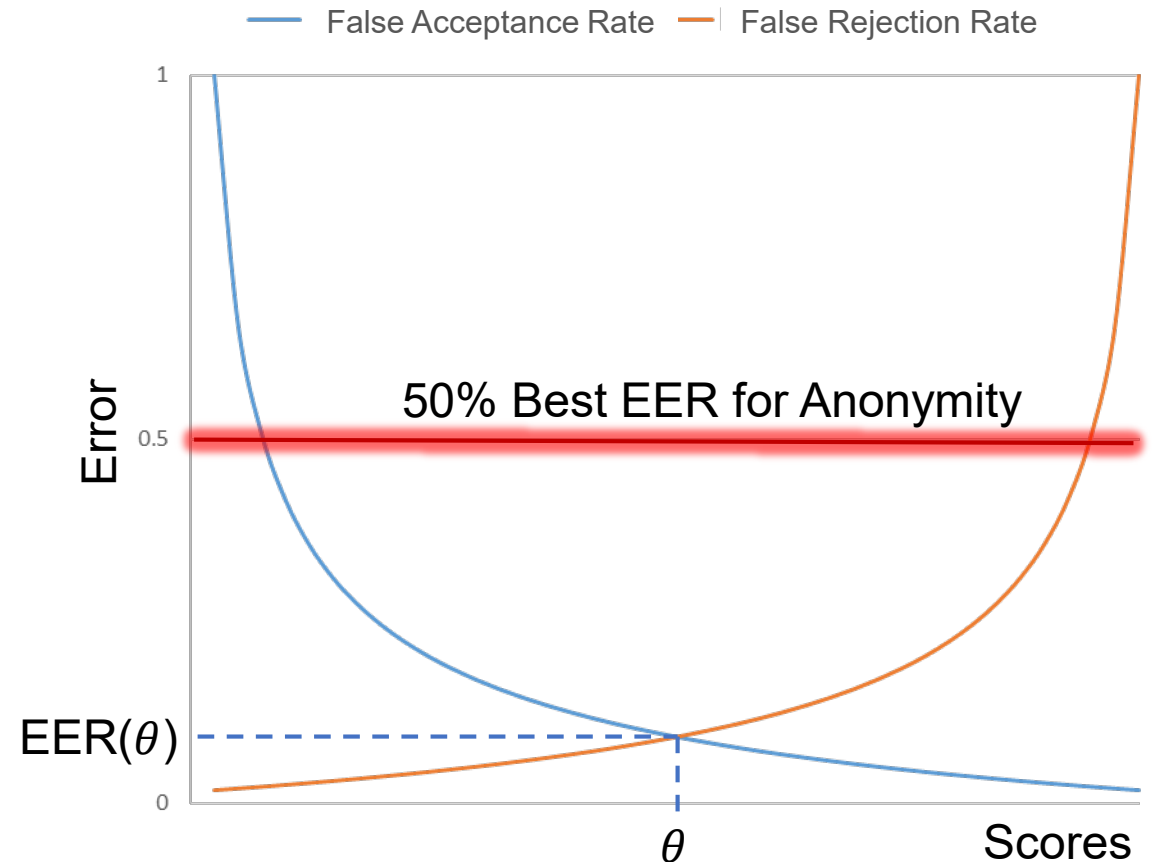
- For ARTS, T&E will be responsible for providing data and product evaluation.

Performer systems will be executed by T&E on stand alone systems

- Specifications of the ARTS Processing System (APS) will be provided at a later date
- ARTS anticipates using multiple T&E teams to evaluate different technical areas and utility constraints.

Consistent with the NIST SRE approach, T&E will use a SID system that compares a test segment against a target, producing a real-valued score  $\theta$ .

- There are no theoretical bounds to  $\theta$ , but larger values mean a more likely match, and lower values a less likely match.
- For a given threshold, we compute errors (false matches, false misses).
- The intersection of these curves is where the errors are equal. EER is the y-value on the curve
- The SID evaluation system will be an x-vector approach with baseline EER of 4.7% on this data.
- The expected EER for non-matching test/target comparisons is 50%.



Panels of listeners will assess static traits of the profile.

- They will first assess the profile from unprocessed data to set a baseline of their ability to differentiate between dialects, ages, and genders.
- The metric will be accuracy for each category, if data sets are balanced:  $\frac{\# \text{ correct predictions}}{\# \text{ of total predictions}}$

Dynamic traits will be assessed through an evaluation of 4k+ low level descriptor (LLD) features derived from the openSMILE 3.0 toolkit.

- These features are widely used for categorization of short-term state, such as emotion, intoxication, sleepiness, health, and have been a key aspect of the INTERSPEECH and ICASSP Challenges.
- T&E will train three deep convolutional neural networks (CNN-4, ResNet-50, and VGG-16) will be LLD features to establish baseline on unprocessed data from the IEMOCAP data set for speech emotion recognition.

Emotions will be categorized as follows:

- Positive: happy, excited, neutral.
- Negative: Anger, sad.

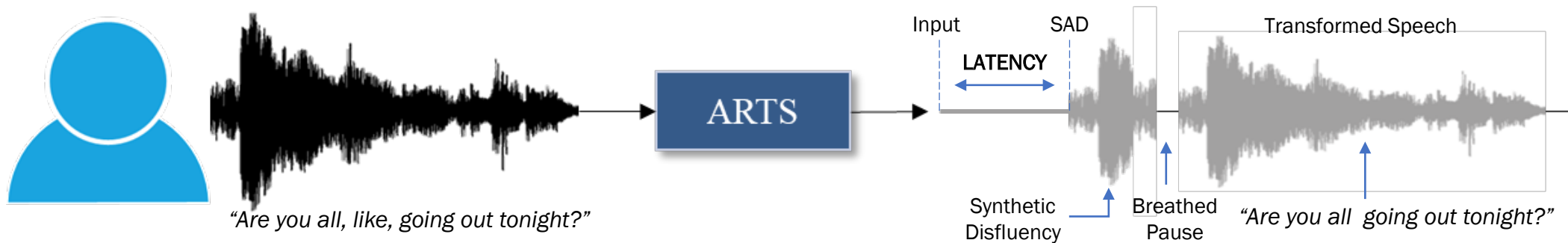
Target metric for classification of transformed is accuracy that reflects a guess.

- Because of imbalances in IEMOCAP dataset, metric will be Unweighted Average Recall (UAR)
- Experimental results have shown UAR of 0.8.

# Utility Constraint Evaluation: Latency

Latency will be measured as the duration between input and speech activity detection.

- All Phase 1 activities will aim for latency on the order of satellite communication.





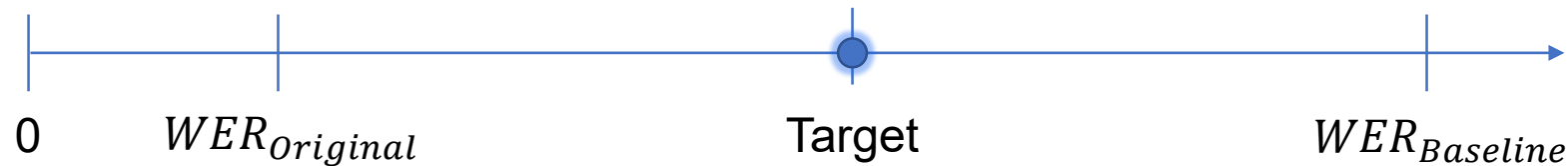
ASR evaluation will be measured by word error rate using Kaldi ASR.

$$WER = \frac{S+D+I}{N}$$

- Discourse markers will be ignored
- $S$  is the number of substitutions (wrong word)
- $D$  is the number of deletions (missing word)
- $I$  is the number of insertions (addition of word)
- $N$  is the number of words

The LDC data used for this evaluation has a WER between 10-12%.

For SID evaluation, the target WER based on the values for original data and baseline data.



The goal in Phase 1 is to decrease this difference by 50%.

- The target metric for ARTS:  $WER = \frac{WER_{Original} + WER_{Baseline}}{2}$

Realism will be assessed using the approach in the Voice Privacy Challenge:

- Listeners are informed they will hear test segments that may be of high quality, but some may sound artificial due to deterioration caused by computer processing.
- Participants evaluate segments on a five-point scale

Rating	Definition	Remark
1	Bad	
2	Poor	Mean score from Voice Privacy Challenge
3	Fair	ARTS Phase 1 target (3.5)
4	Good	ARTS Phase 2 target (4.5)
5	Excellent	

The panel will consist of ~200 unique listeners, with approximately equal number of male/female, each evaluating an average of 5 sets/participant

- These figures exceed but are of similar magnitude of the number of participants used for the Voice Privacy Challenge.



# Phase 1 Target Metric Summary



Measurement	Metric	Target	Description of Target
SID	Equal Error Rates	45%	EER at 50% is expected value for comparison of non-matching segments.
Static Traits	Accuracy	$Acc_{Original}$	Best ability to differentiate between classes with original, unprocessed data.
Dynamic Traits	UAR	48%	Classifiers produce output that equals random guessing.
Latency	Average duration from input to speech detection	350ms	Communication on the order of satellite channels.
Understandability	Average word error rate based on Kaldi STT	$\frac{WER_{Original} + WER_{Baseline}}{2}$	Drives WER towards a target similar to understandability of original data.
Naturalness	Mean opinion score	3.5	Reflects average assessment of fair to good naturalness.



# ARTS Target Metric Summary



Measurement	Metric	Phase 1 (English) Target	Phase 2 (Multi-Lingual) Target
<b>SID</b>	Average Equal Error Rates	45%	45%
<b>Static Traits</b>	Accuracy	$Acc_{Original}$	$Acc_{Original}$
<b>Dynamic Traits</b>	UAR	48%	48%
<b>Latency</b>	Average duration from input to speech detection	350ms	150ms
<b>Understandability</b>	Average word error rate based on Kaldi STT	$\frac{WER_{Original} + WER_{Baseline}}{2}$	$\frac{3 * WER_{Original} + WER_{Baseline}}{4}$
<b>Naturalness</b>	Mean opinion score	3.5	4.5

Feedback, thoughts and comments:

- ARTS Team Alias: [dni-iarpa-ARTS-proposers-day@iarpa.gov](mailto:dni-iarpa-ARTS-proposers-day@iarpa.gov)

Additional information:

- ARTS website: [www.iarpa.gov/index.php/research-programs/ARTS](http://www.iarpa.gov/index.php/research-programs/ARTS).

